Algorithm and Data Fusion to Improve Estimates of Forest Status and Change

(and some other good stuff)

Randolph H. Wynne and Valerie A. Thomas

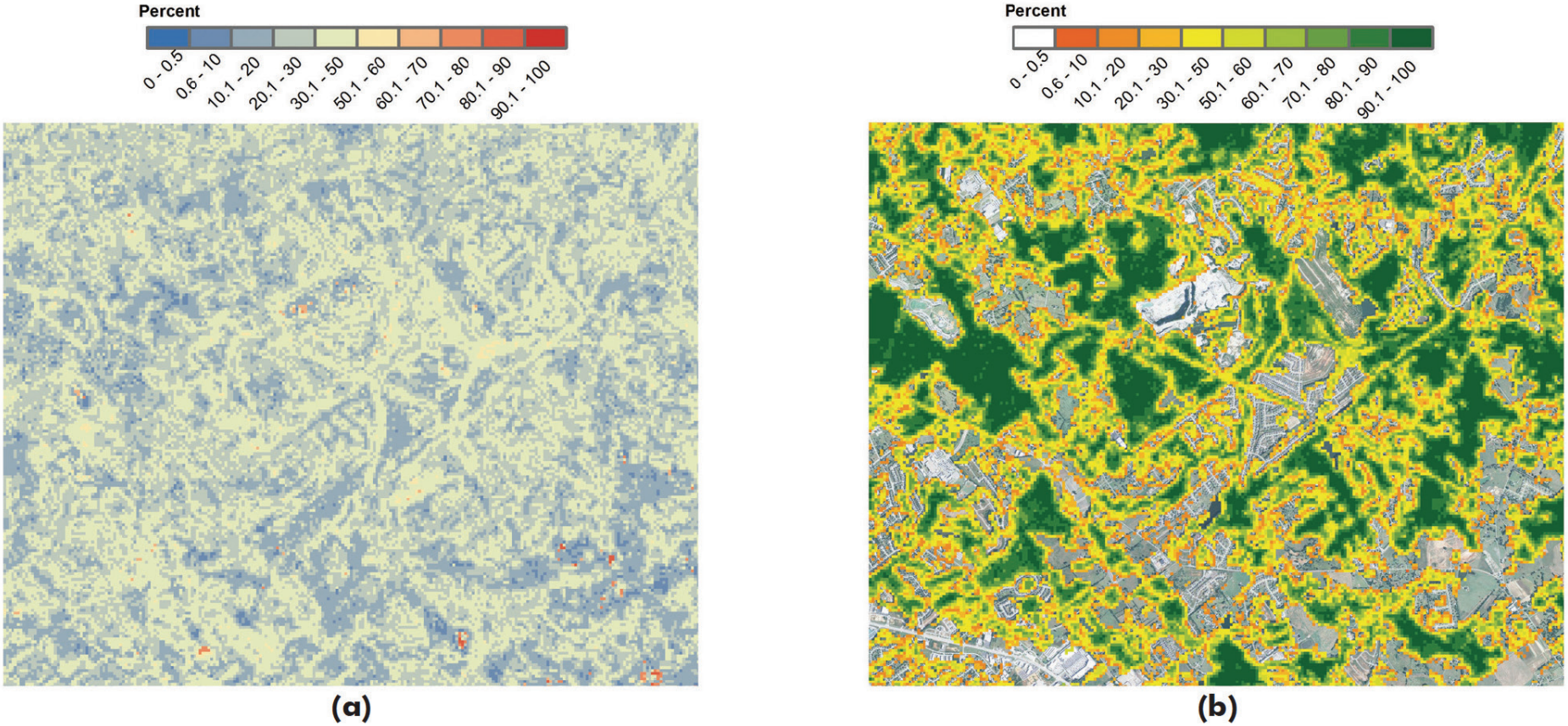
C.E. Blinn, E.B. Brooks, J.O. Coulston, J.M. Derwin, T.R. Fox, S. Ghannam, M.N. House, K. Moeltner, S.S. Peery, R. Saxena, L.T. Watson, L. Yu

Take Homes

- Probability underutilized
- Crowds good for more than clouds
- Time series rock
- 900 m^2 often too big
- Pixels have neighbors
- Algorithms and computation essential to our science

Approximating Prediction Uncertainty for Random Forest Regression Models

John W. Coulston , Christine E. Blinn , Valerie A. Thomas , and Randolph H. Wynne

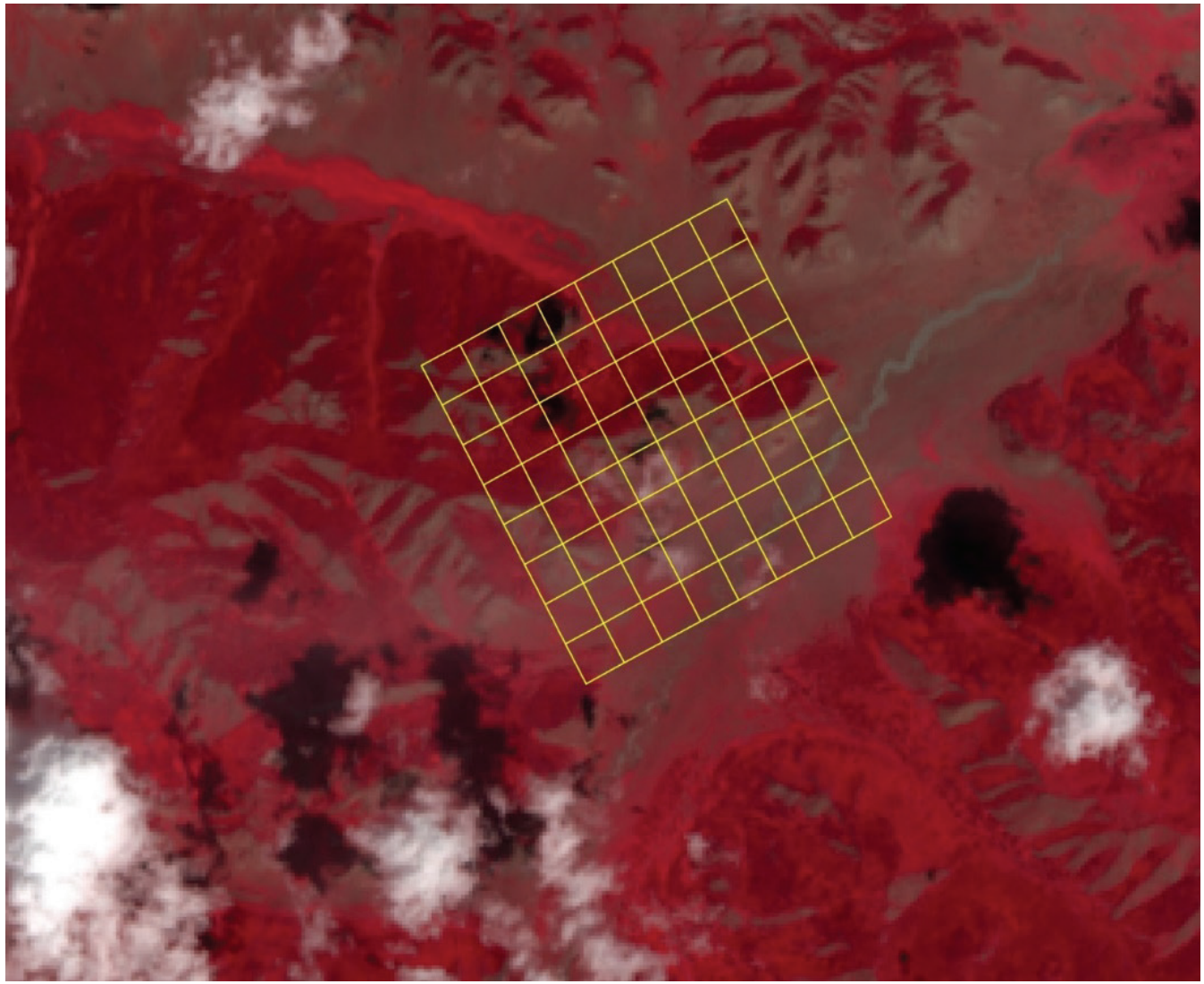


(a) (b)

Plate 4. Half-width of the 95 percent prediction interval for percent tree canopy cover for a portion of (A) the Georgia study area, and masked predicted percent tree canopy cover with 5 percent error rate for area of no canopy cover overlaid on (B) the NAIP imagery.

Cloud Sourcing

Image 5



Heatmaps: Image 5

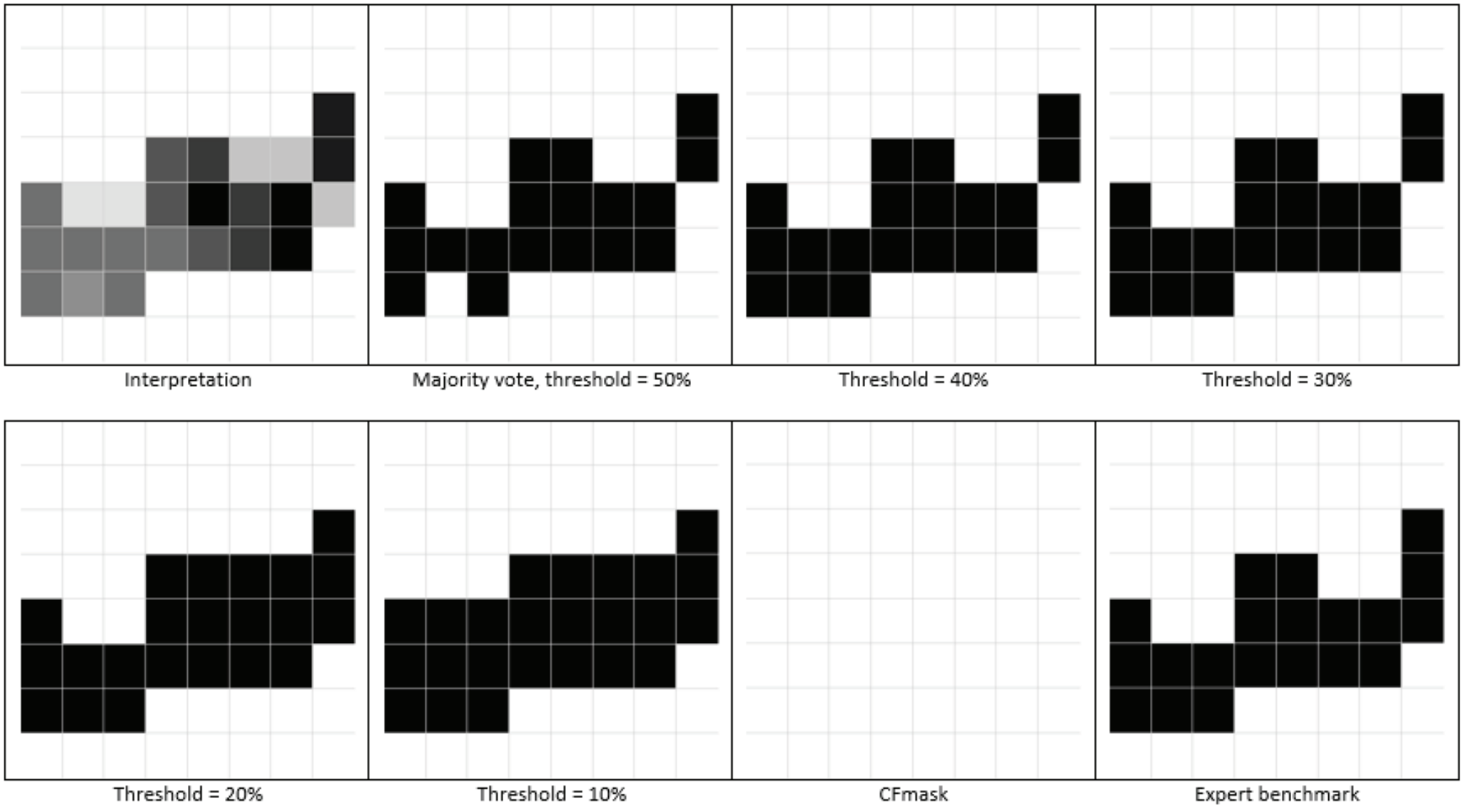
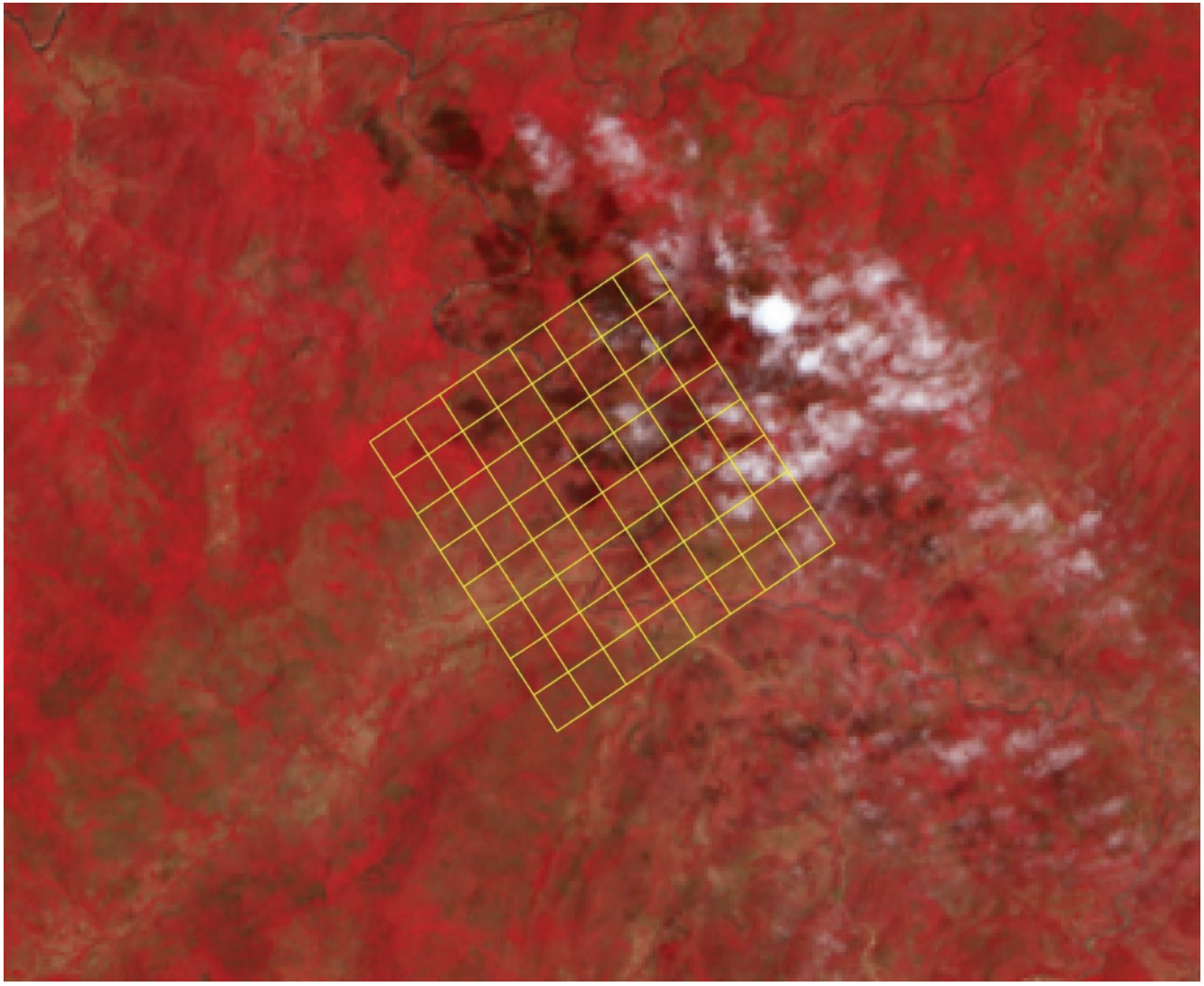
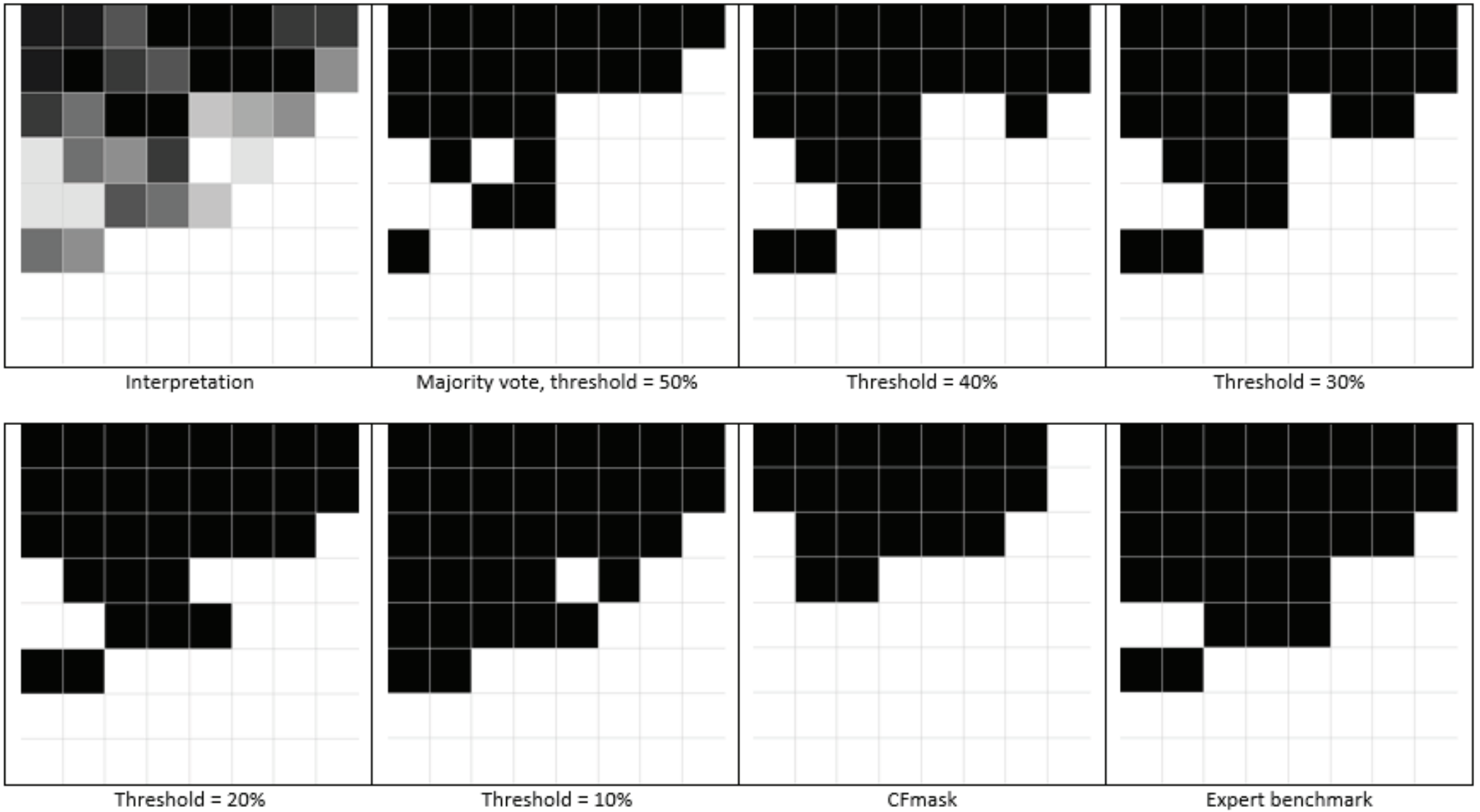


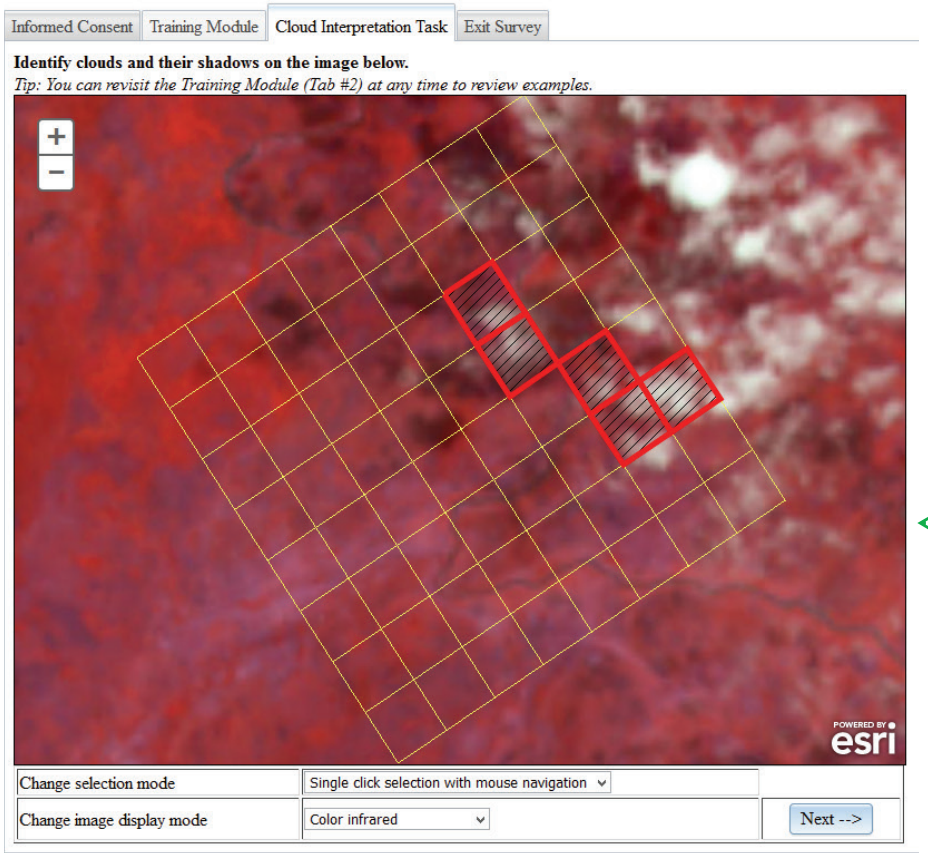
Image 6



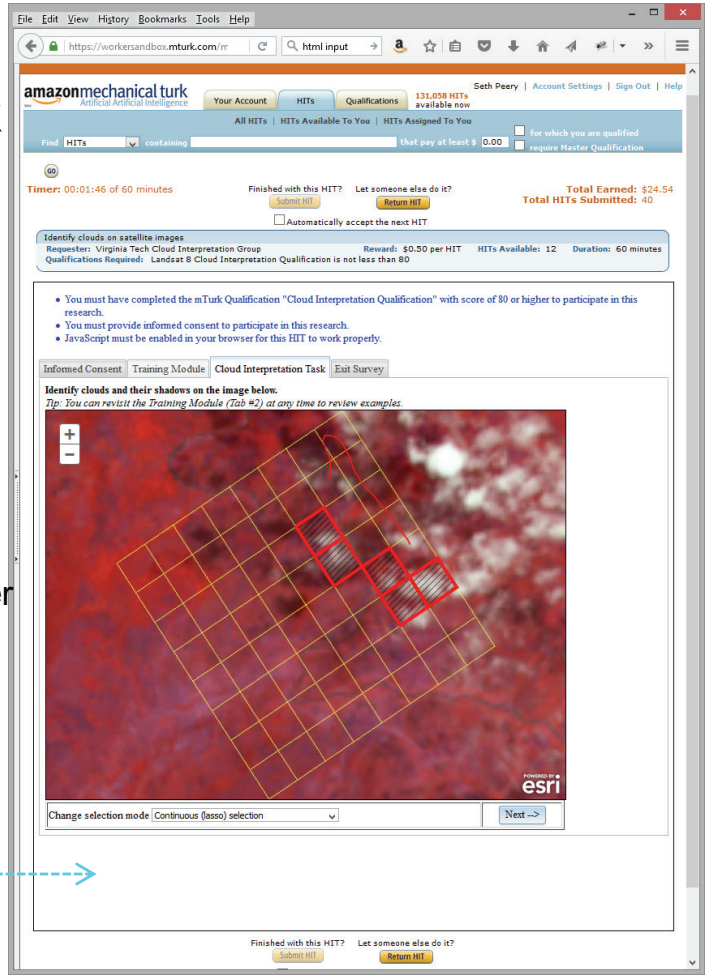
Heatmaps: Image 6



Cloud Sourcing ”Machinery”
(crouds good for more than clouds)



VT←→mTurk



ExternalQuestion

ESRI ArcGIS Server
JS API App
located at VT

p092r086_c1 : 2
p092r086_c2 : 2
p092r086_c3 : 2
p092r086_c4 : 0
p092r086_c5 : 0
p092r086_c6 : 0
p092r086_c7 : 0
p092r086_c8 : 0

p092r086_b1 : 2
p092r086_b2 : 2
p092r086_b3 : 2
p092r086_b4 : 0
p092r086_b5 : 0
p092r086_b6 : 0
p092r086_b7 : 0
p092r086_b8 : 0

p092r086_a1 : 2
p092r086_a2 : 2
p092r086_a3 : 2

Cloud impacted tiles are returned to mTurk, as {ID:interp code}

Response Variable: Photo Interpreted TCC%

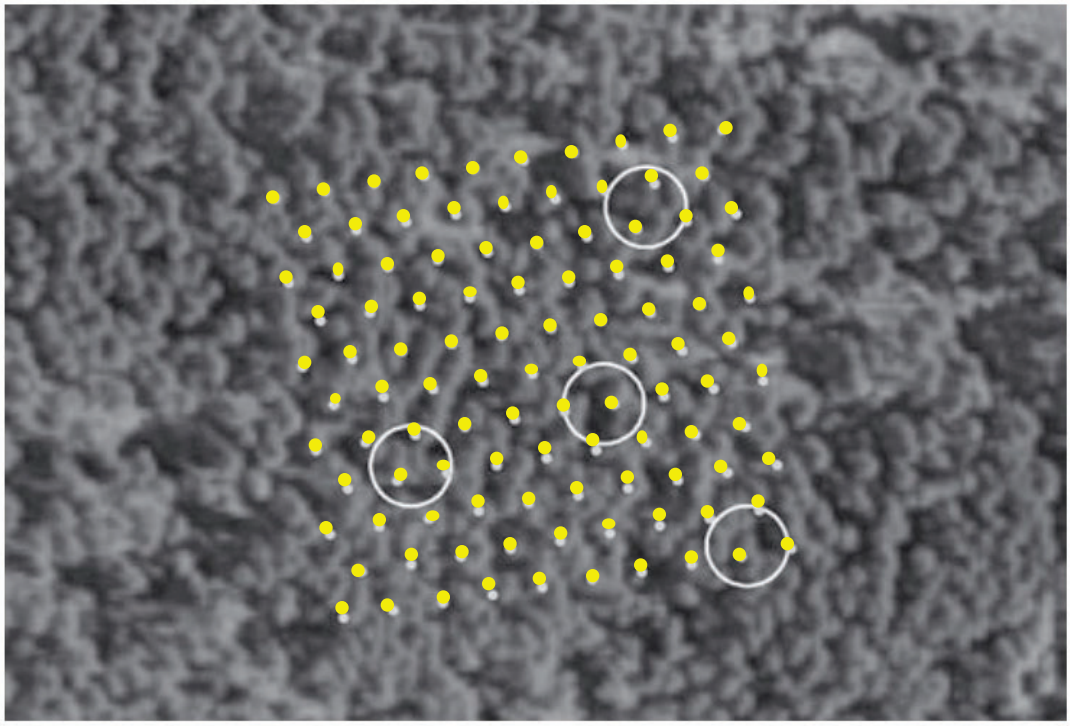


Photo Interpretation Technique, image from Toney, 2010

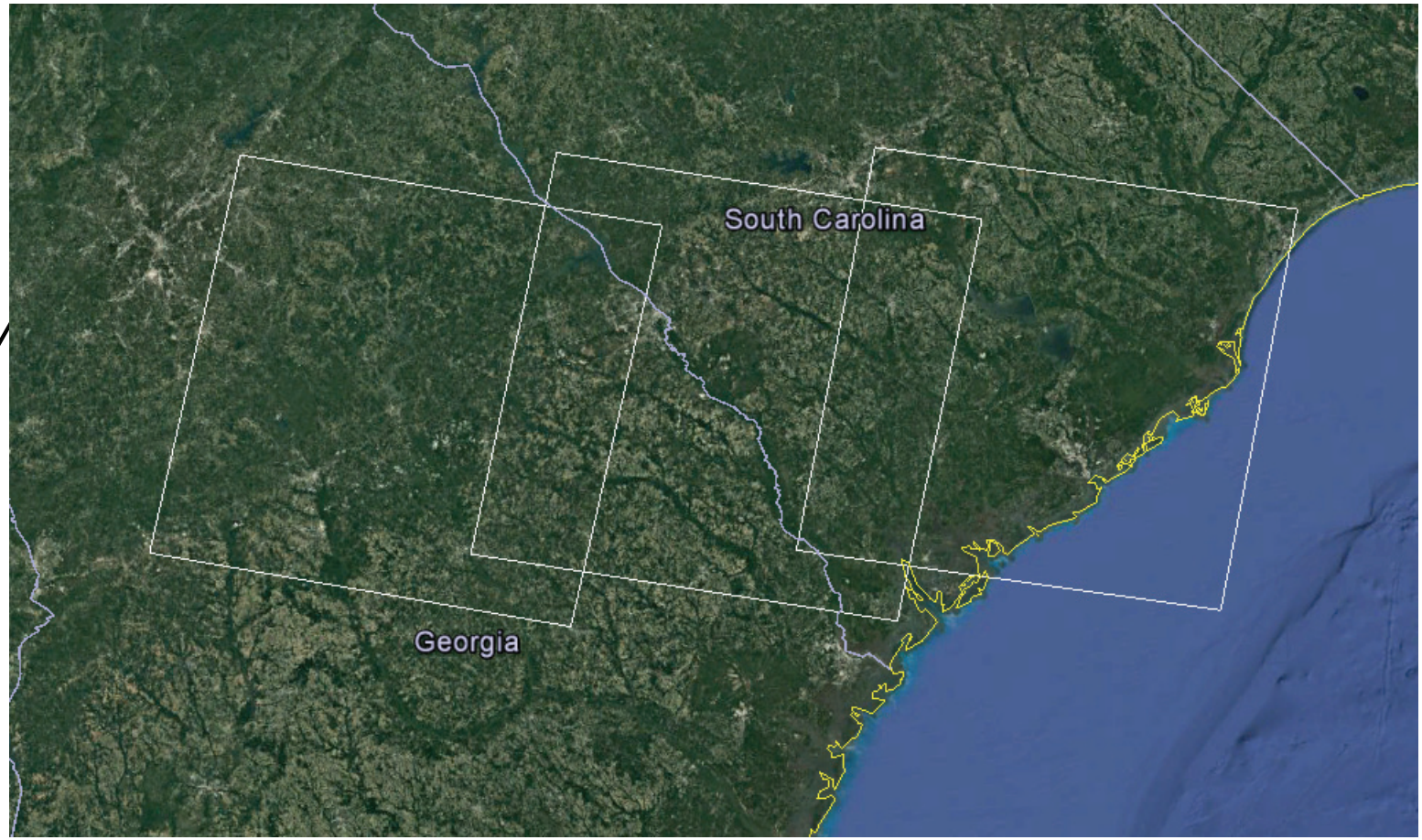
- 105 photo points per sampling location
- Overlaid on NAIP imagery
- 90 m^2 area

Study Area, ‘West’

WRS path/row 43/30, 44/30, 45/30 in OR

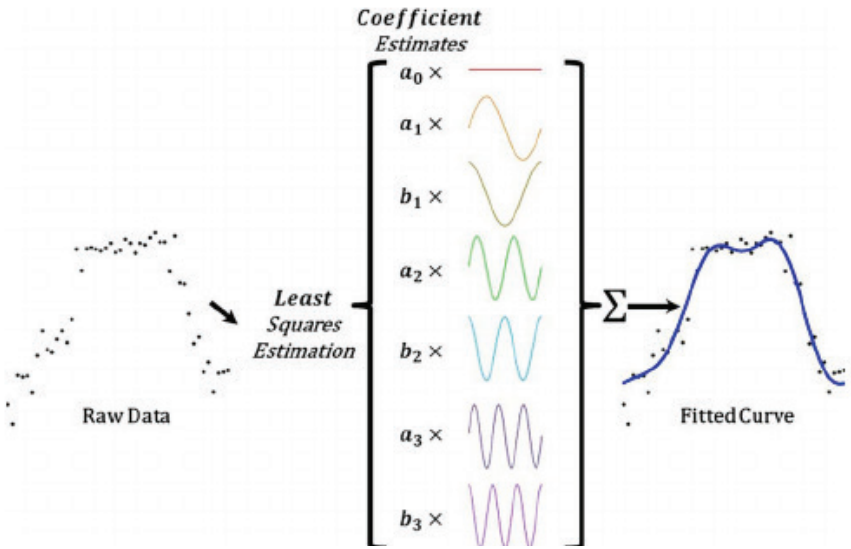
Study Area, 'South'

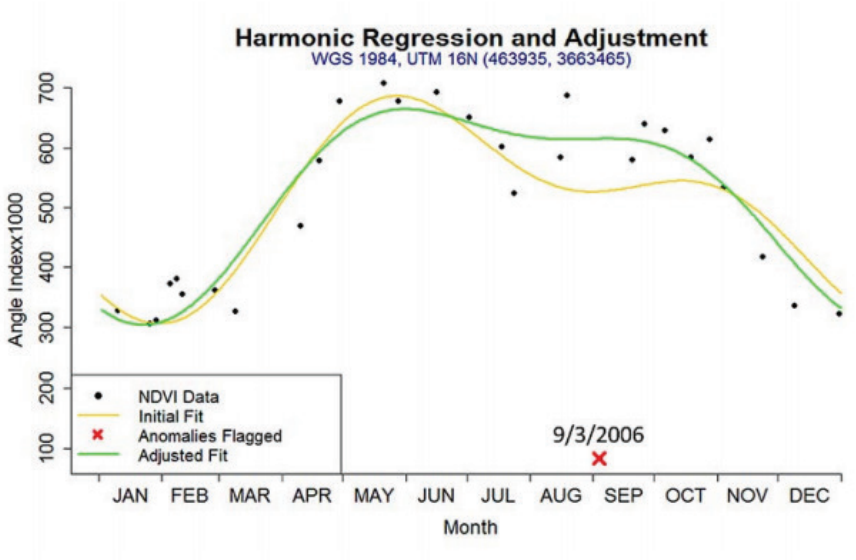
WRS path/row 16/37,17/37,18/37 in GA and SC

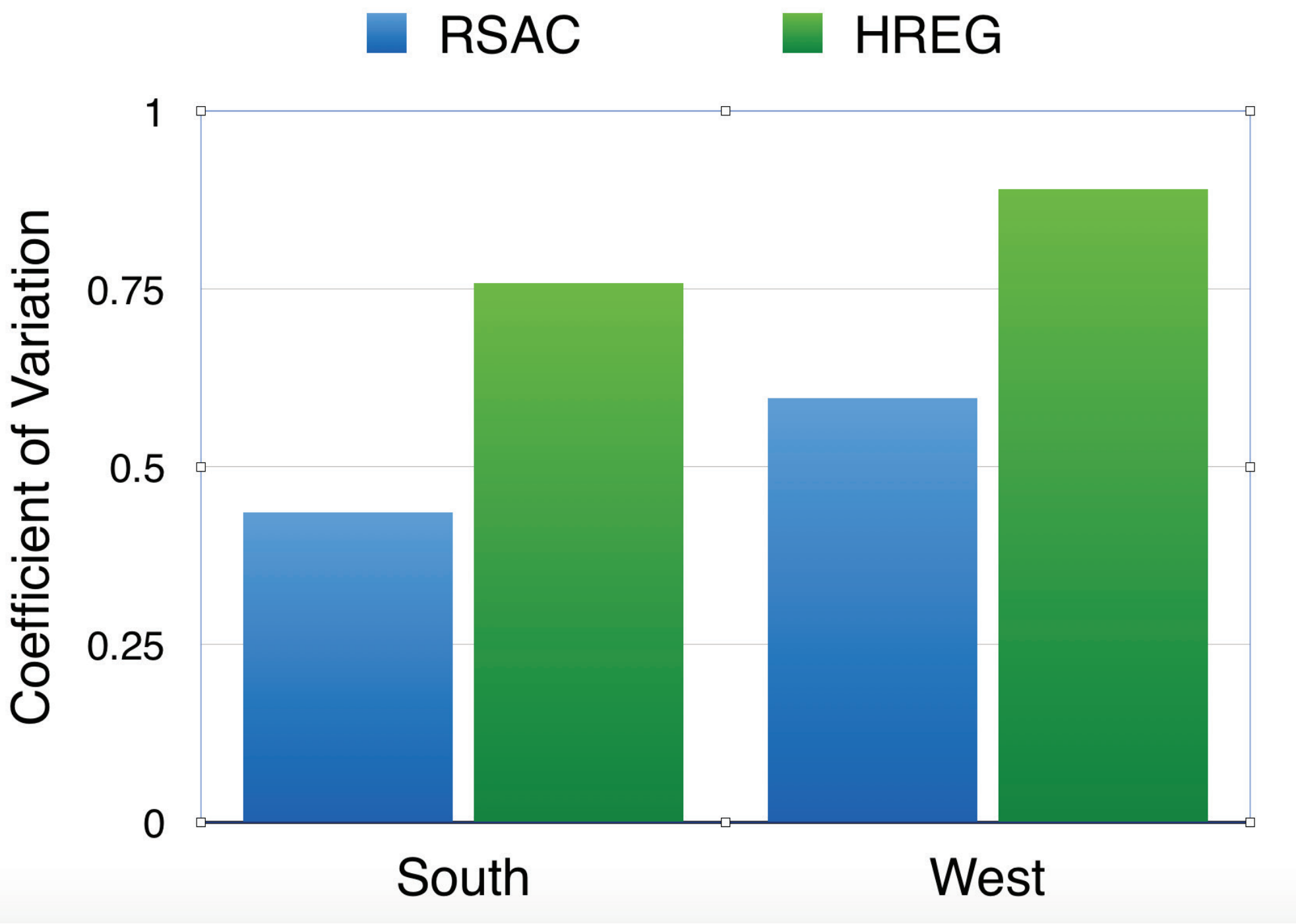


Multi-Date, Multi-Band Time Series Variables

- NDVI, SWIR 5, SWIR 7 images from 1984-2014
- 3 yrs around time of TCC estimate
- "Five number summary": constant, 2 cosine coefficients, 2 sine coefficients



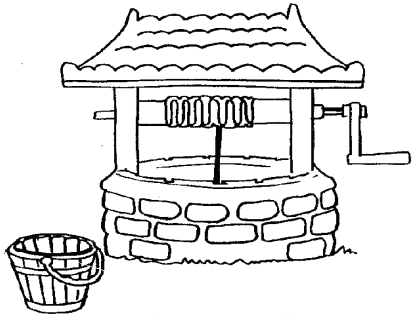



RSAC
HREG
Coefficient of Variation
1
0.75
0.5
0.25
0
South
West

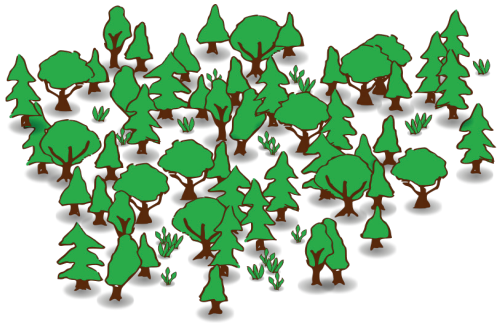
Mapping Low-Density Exurban Development using Landsat

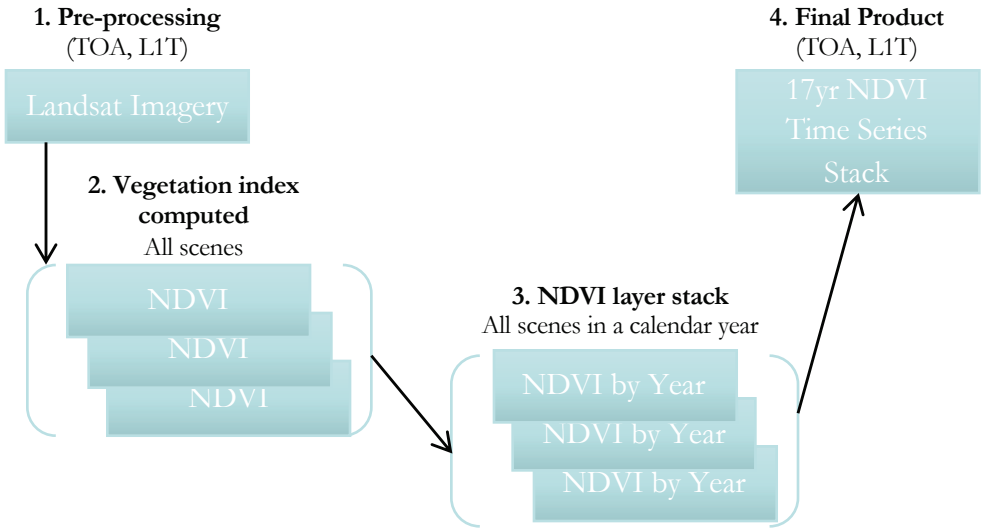
Housing Starts / Permits

Well Permits



1992 NLCD Forest Layer

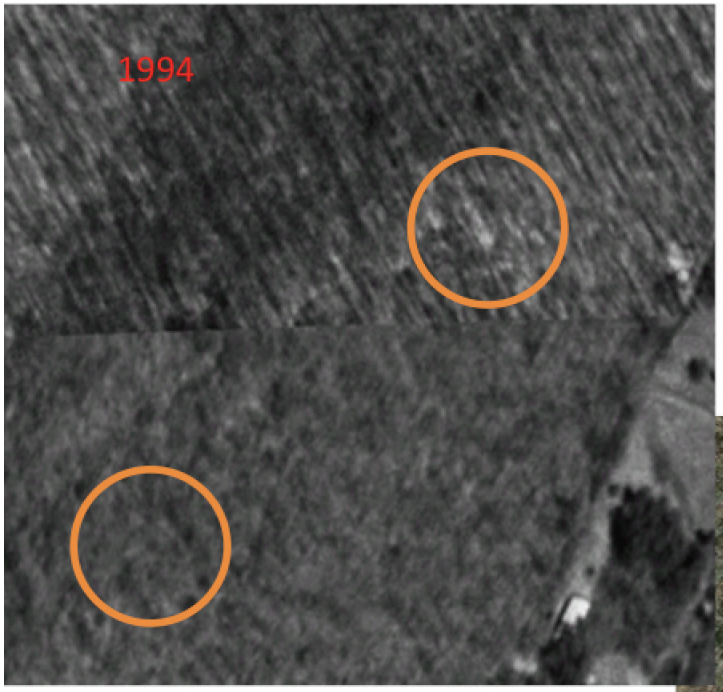




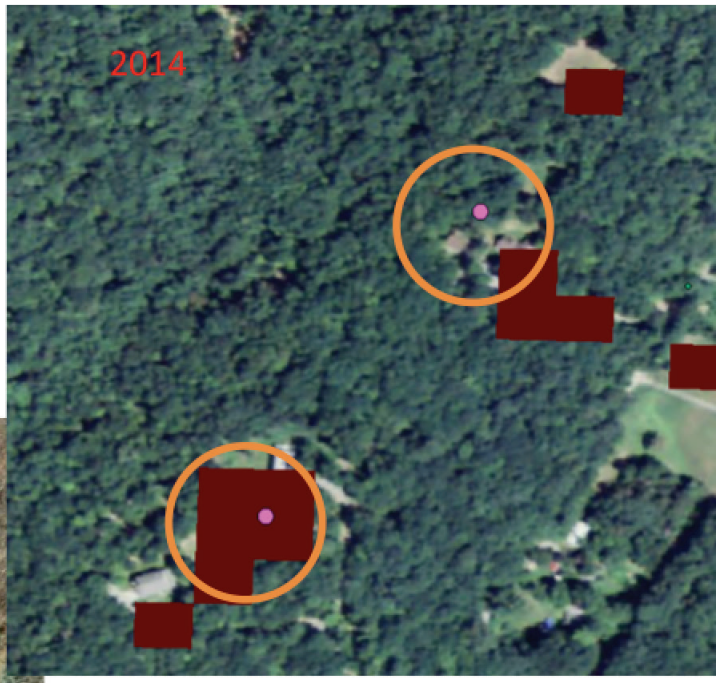
Program Variables

1. Slope of recovery
2. Slope of 3 yrs after low
3. Low (value when first fell below threshold)
4. High (highest value after Low)
5. Average of 3 yrs after Low
6. Average of lowest 3 values in time series

Hotspot Conversion Tracking Through Landsat



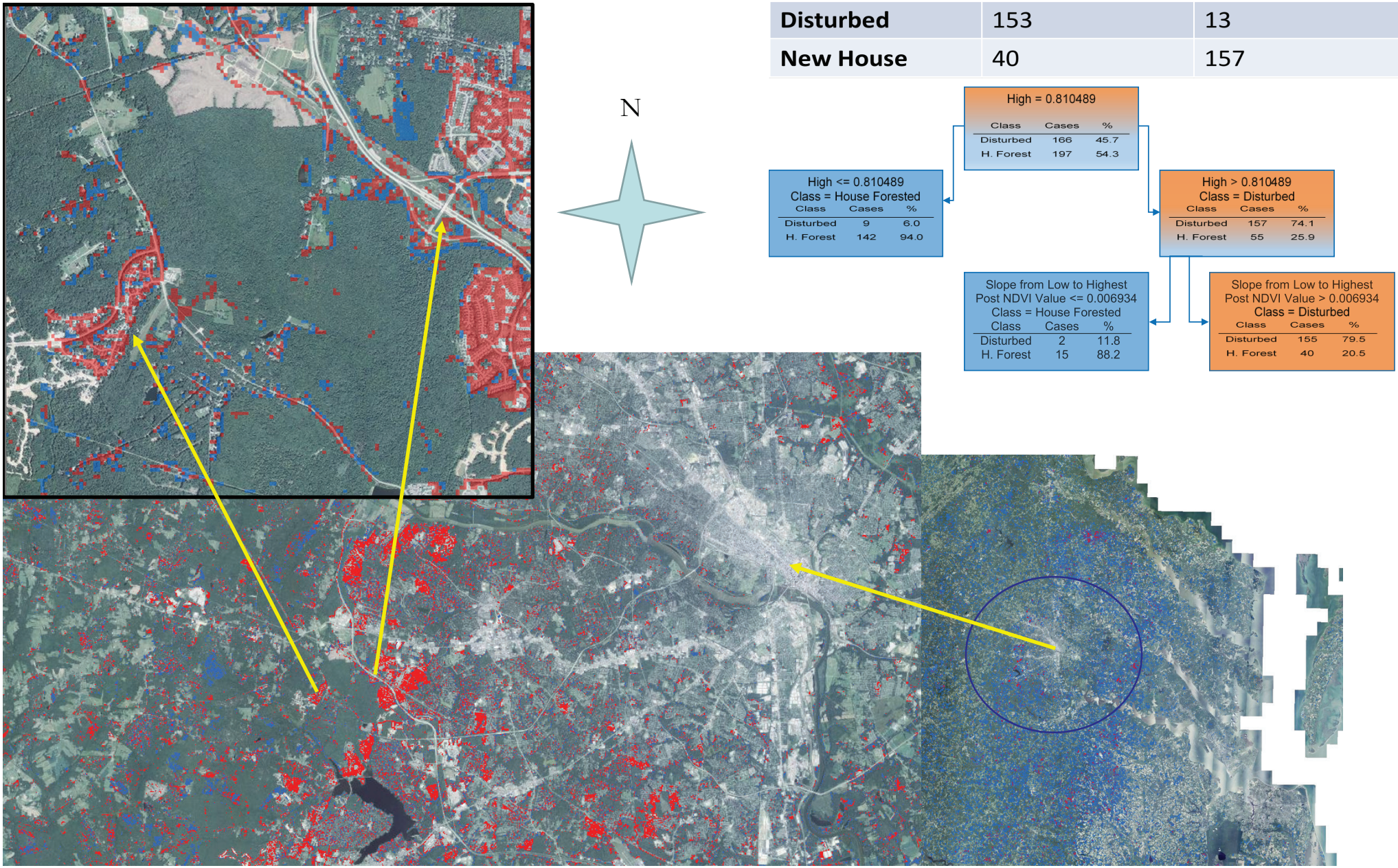
Using housing permit data we are able to verify that the program can accurately detect subtle disturbances to the forest ecosystem.



900 m^2 pixel often too big

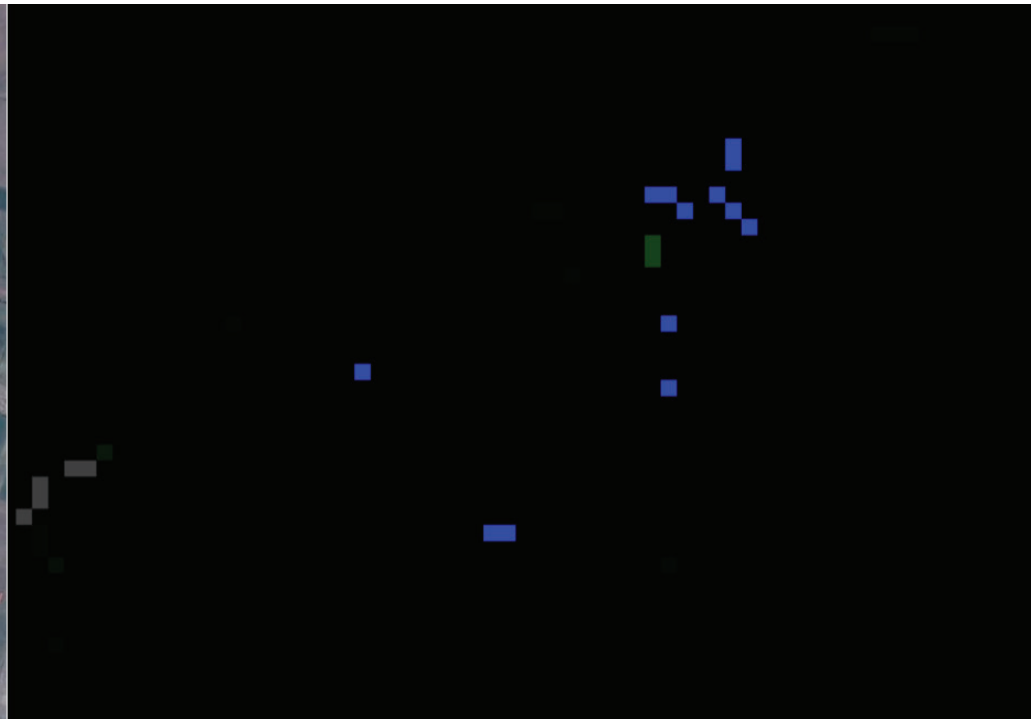
1332 individual scenes across the state of Virginia, ranging from 1995-2011 and spanning 13 different Landsat path/rows

	Disturbed	New House
Disturbed	153	13
New House	40	157



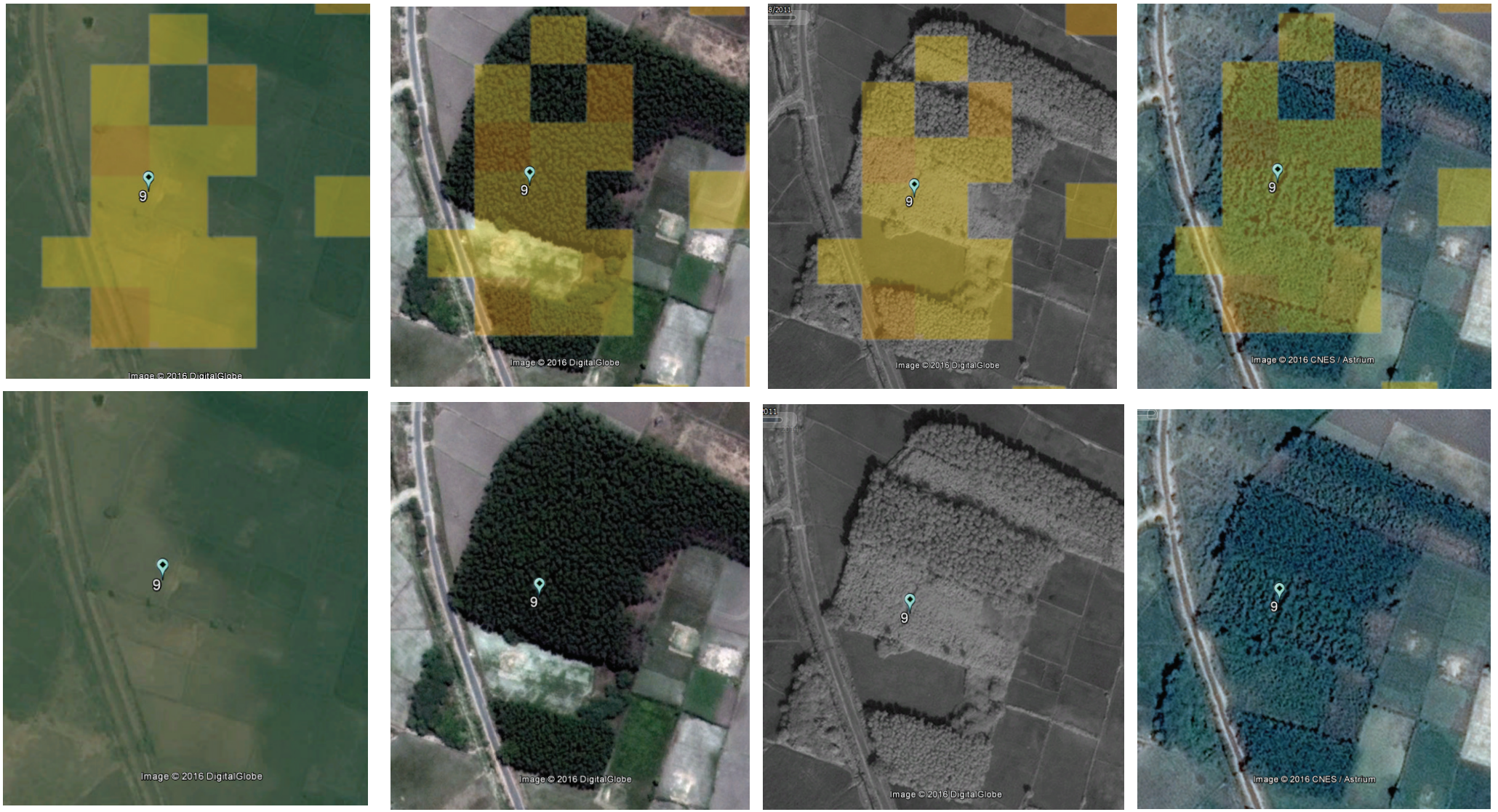
“Trees outside forests” important for carbon exchange and forest-based economies but often at finer scale than Landsat can comfortably handle

SPOT image courtesy Google Earth showing newly established small plantations in Andhra Pradesh

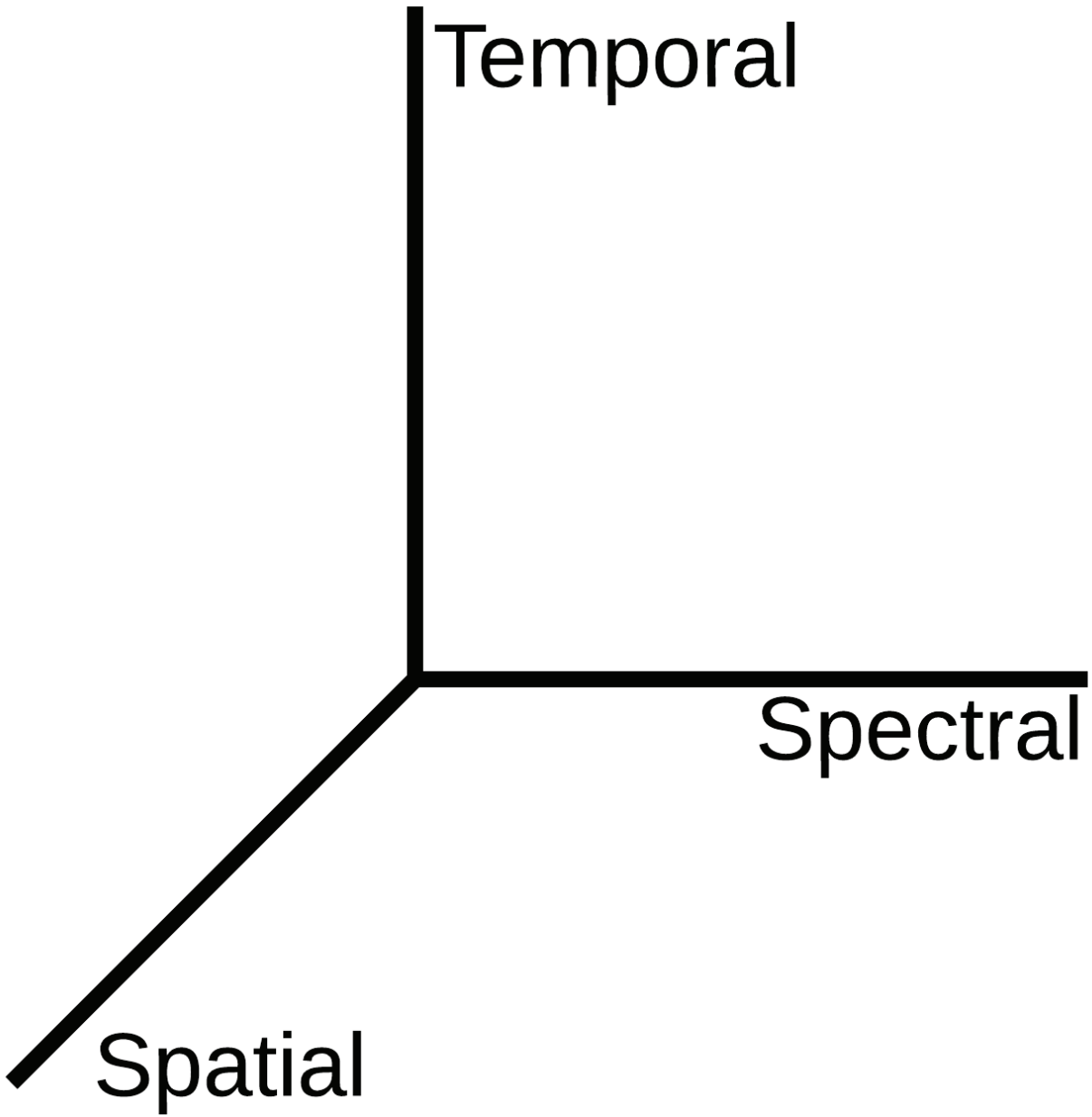


Hansen et al. (2013) from same period with blue showing forest gain and green prior forest extent

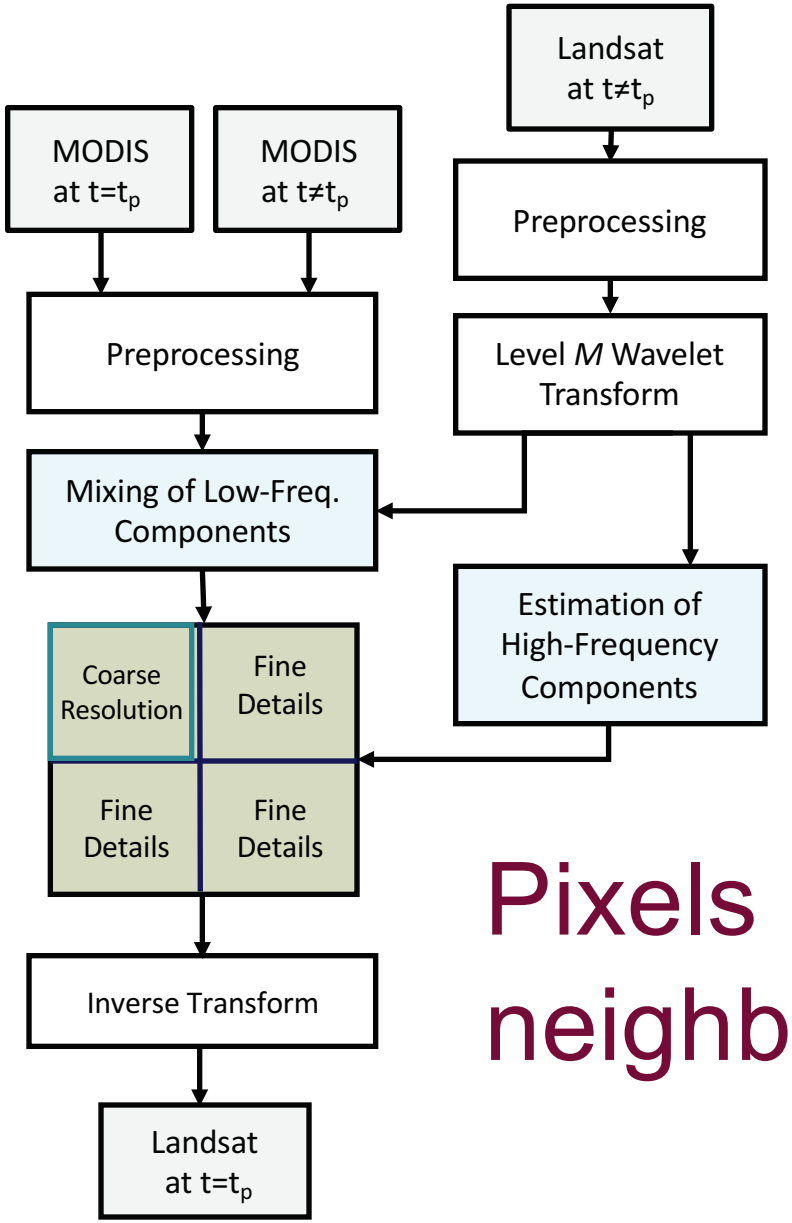
Time series help, but more spatial resolution needed



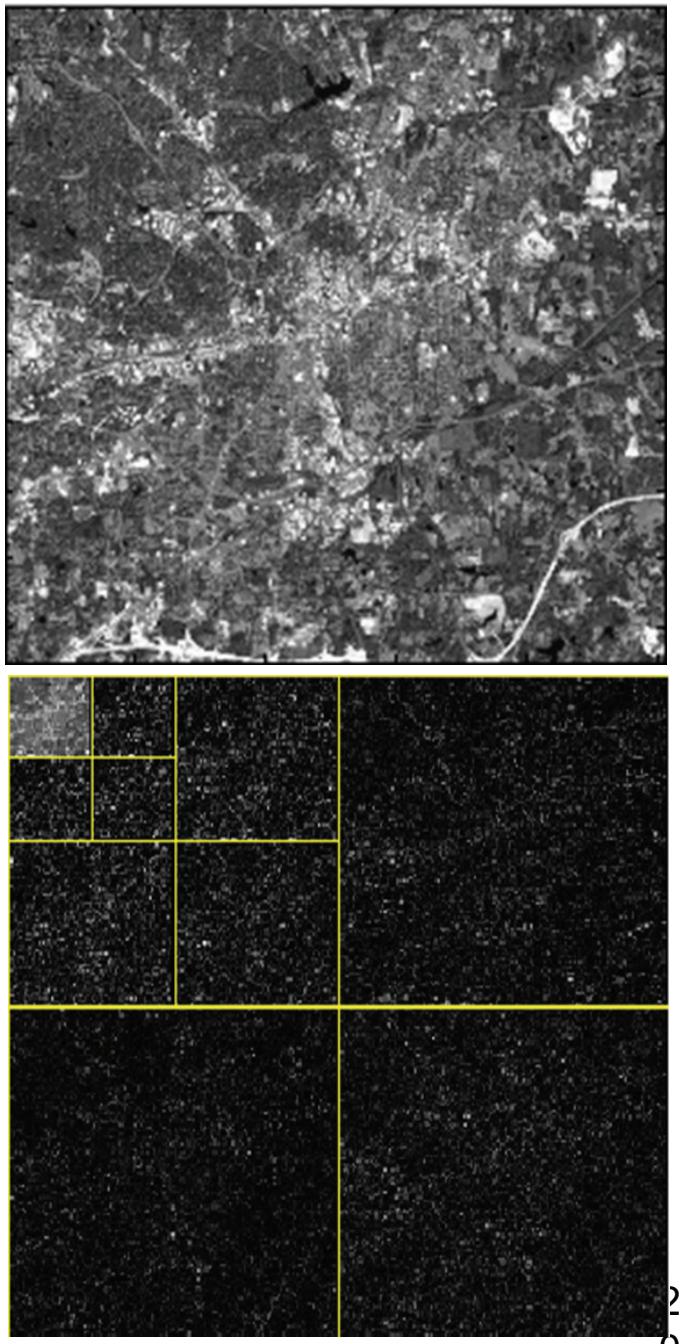
EWMACD (SWIR2, 1 harmonic) classification of plantations Andhra Pradesh. 166 Landsat 7 and 8 images from path 143 row 49 spanning from 2000 to 2016 were used in this analysis.


Temporal
Spectral
Spatial

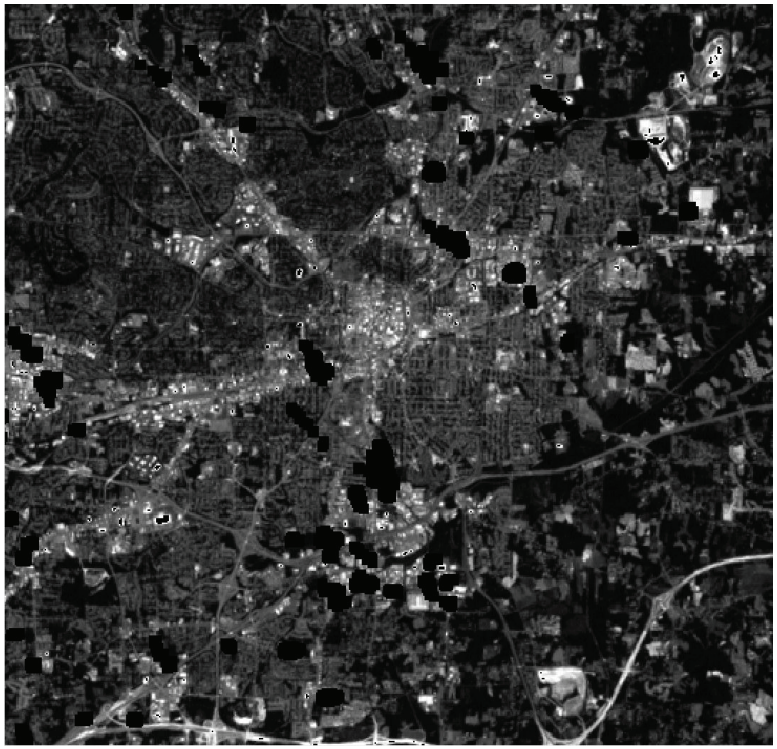
Wavelet-based Spatiotemporal Adaptive Data Fusion Model (WSAD-FM)



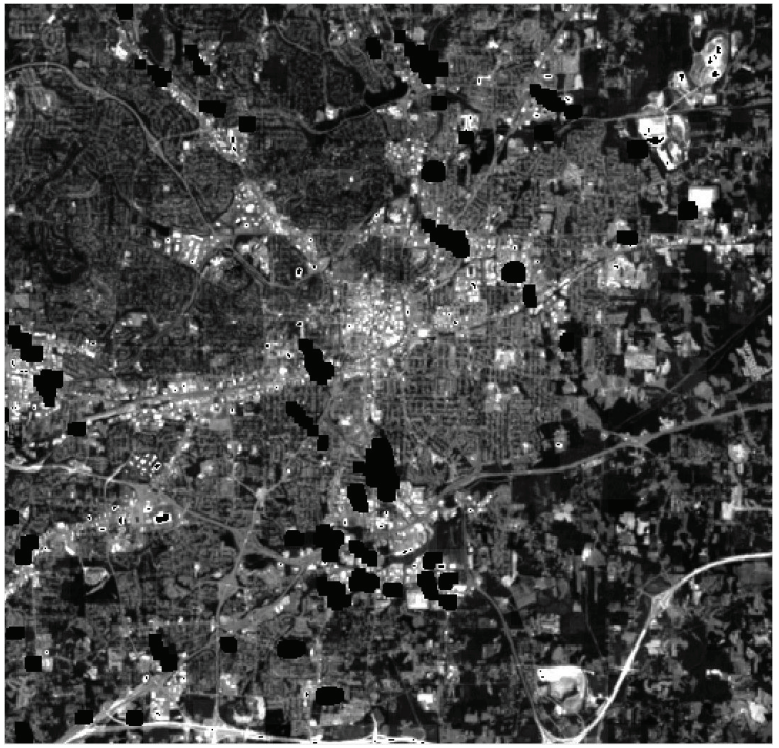
Pixels have neighbors



Results: May 24, 2002



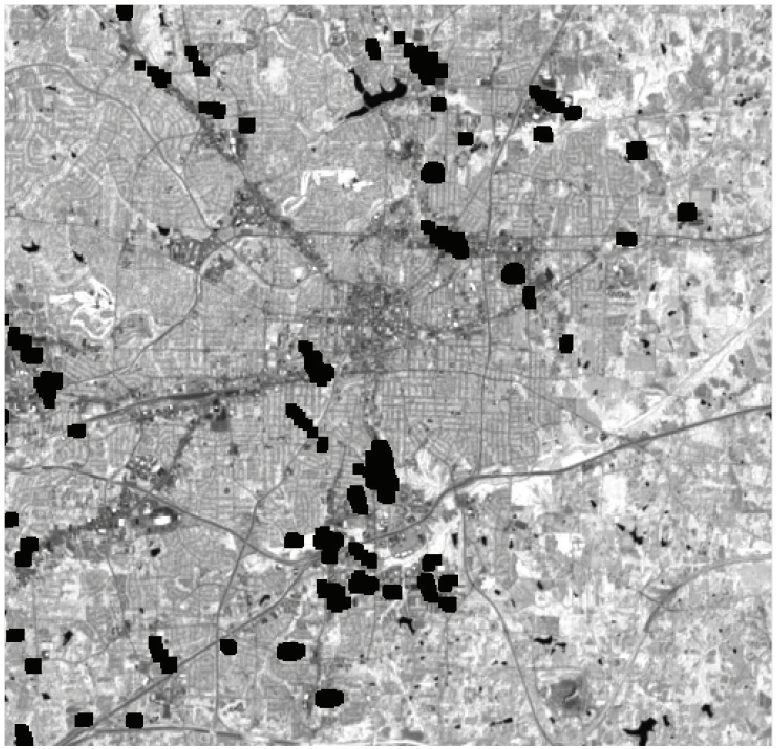
Actual red band image



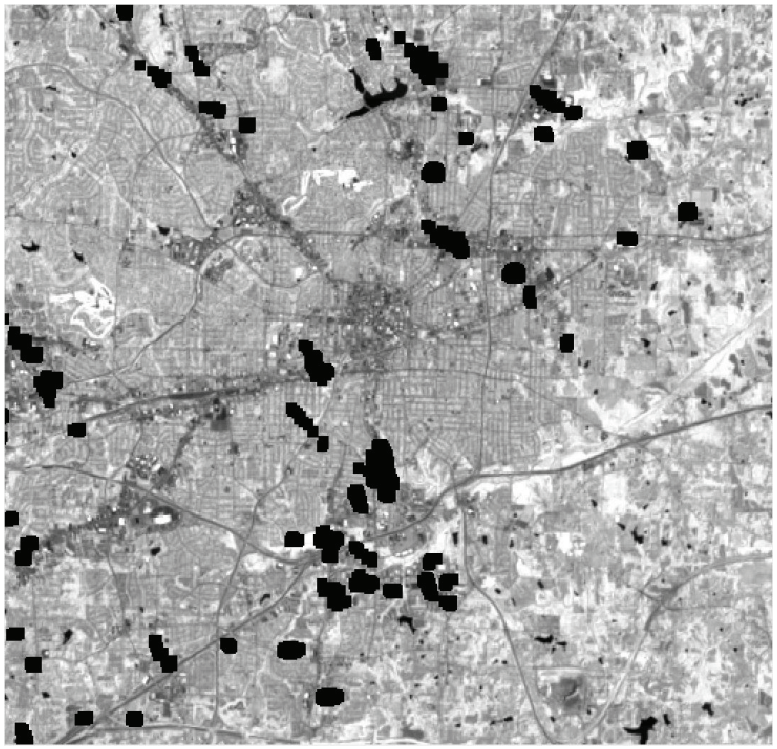
Predicted red band image

$R^2 = 0.9450$

Results: May 24, 2002

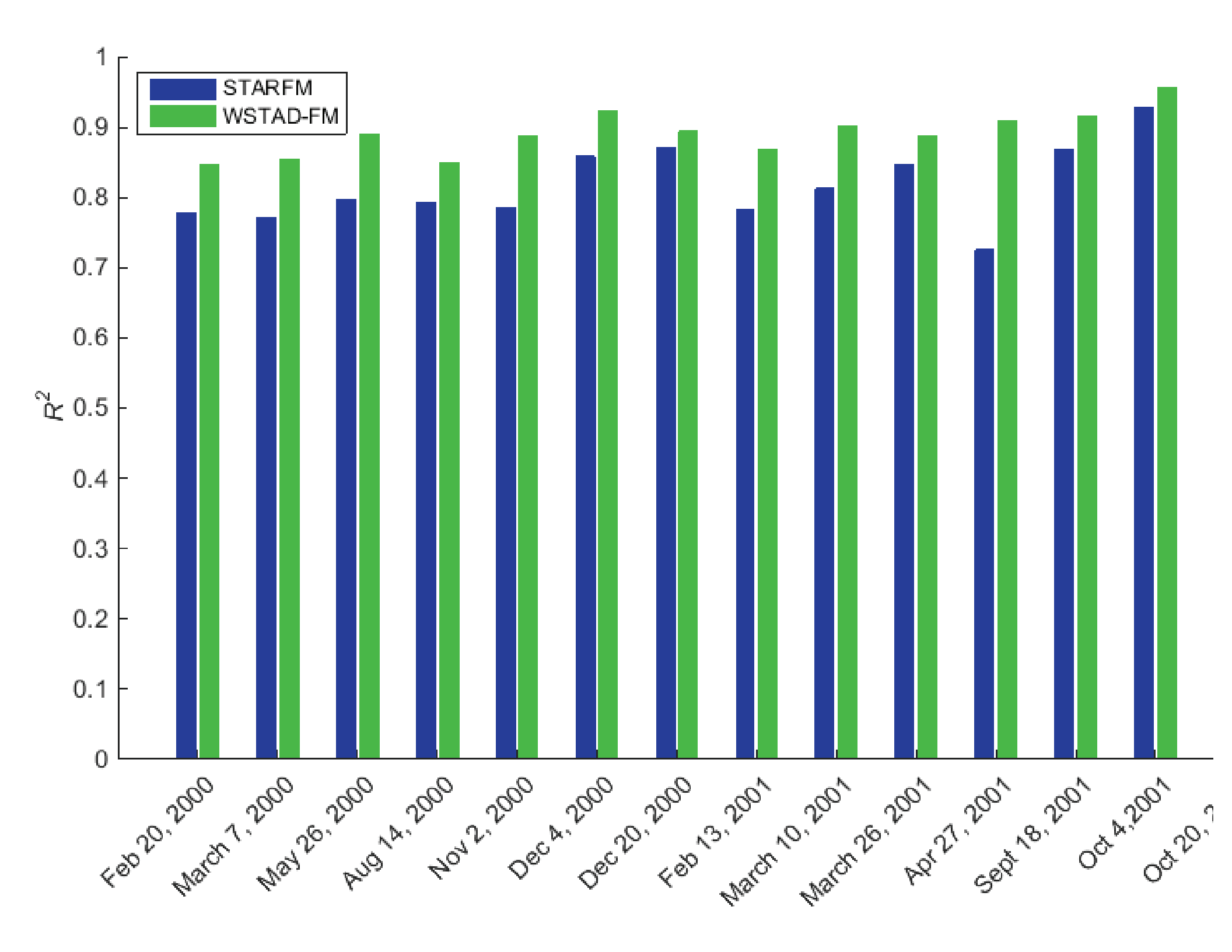


Actual near-infrared band image



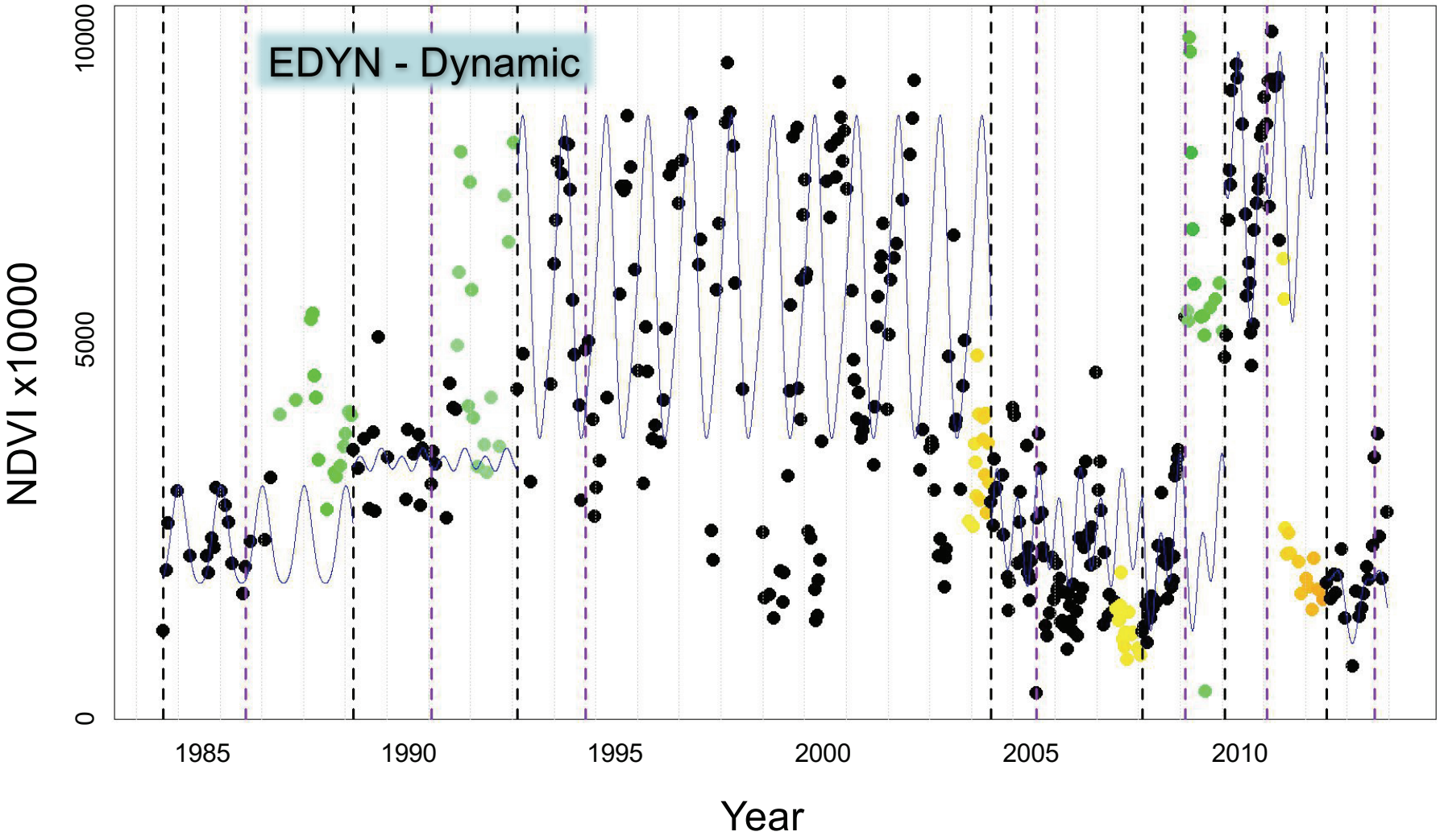
Predicted near-infrared band image

$R^2 = 0.9126$


STARFM
WSTAD-FM
R^2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Feb 20, 2000
March 7, 2000
May 26, 2000
Aug 14, 2000
Nov 2, 2000
Dec 4, 2000
Dec 20, 2000
Feb 13, 2001
March 10, 2001
March 26, 2001
Apr 27, 2001
Sept 18, 2001
Oct 4,2001
Oct 20, ?

We are scientists who compute!

EWMACD/Updates



LandTrendR

Notation and definitions: For an $m \times n$ matrix A, an n-vector x, $I \subset \{1, \ldots, m\}$, $J \subset \{1, \ldots, n\}$, let A_{IJ} denote the submatrix of A formed from the rows indexed by I and the columns indexed by J, and x_J denote the subvector of x indexed by J. $A_{I\cdot}$ ($A_{\cdot J}$) are the rows (columns) of A indexed by I (J), respectively. An image is an $R \times C$ matrix D, where each D_{rc} (pixel) is an $S \times B$ matrix, whose (s, b) element $(D_{rc})_{sb}$ is the signal value at time index s and frequency band index b.

Algorithm LandTrendR.

for band $b = 1 : B$
  for row $r = 1 : R$
    for col $c = 1 : C$ **do**

Step 1: Despike

Let $u = (D^0_{rc})_{\cdot b}$ denote the raw time series data. For each time point t_i, $1 < i < S$, define $\Delta u_i = (D^0_{rc})_{(i+1)b} - (D^0_{rc})_{ib}$, $\nabla u_i = (D^0_{rc})_{ib} - (D^0_{rc})_{(i-1)b}$, $\mu\delta u_i = (D^0_{rc})_{(i+1)b} - (D^0_{rc})_{(i-1)b}$, $k_i = 1 - |\mu\delta u_i| / \max\{|\nabla u_i|, |\Delta u_i|\}$, and correction

$$\kappa_i = (\delta^2 u_i) k_i / 2 = \big((D^0_{rc})_{(i-1)b} - 2(D^0_{rc})_{ib} + (D^0_{rc})_{(i+1)}\big) k_i / 2.$$

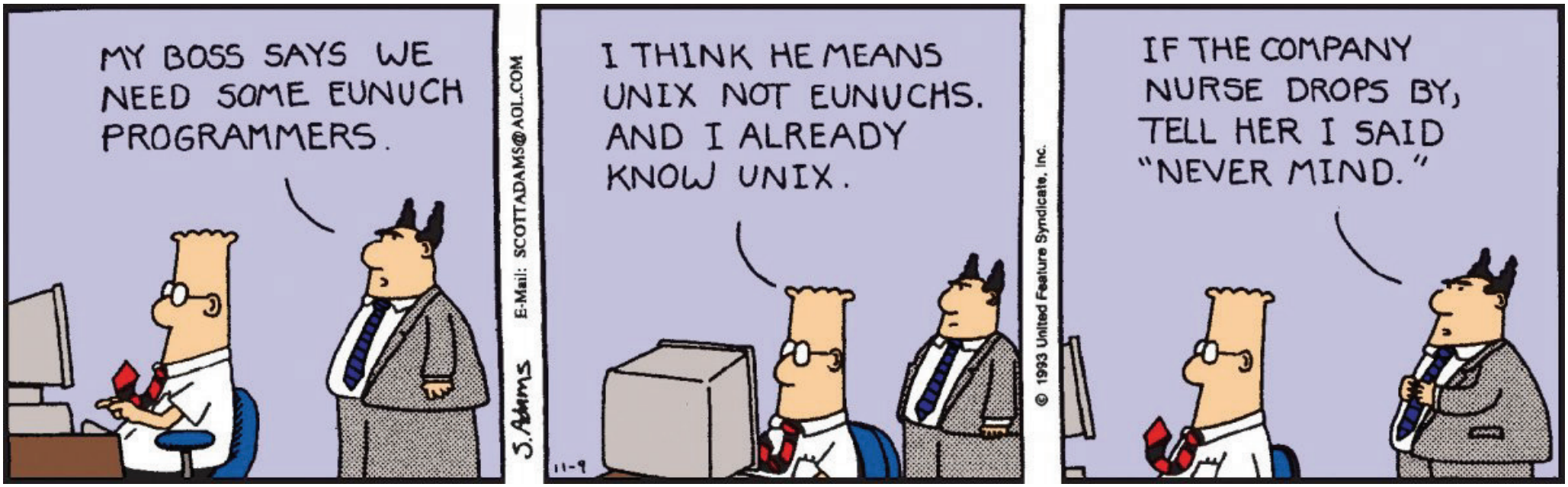
For each i such that $k_i = \max\limits_{1<j<S} k_j$, update $(D_{rc})_{ib} := (D^0_{rc})_{ib} + \kappa_i$. Repeat iteratively until $\max\limits_{1<j<S} k_j < \upsilon$, some given despiking tolerance.

Step 2: Find potential breakpoints

Let $S^1 = (t_1, \ldots, t_S)$ be the original sequence of time points and $I^1 = (2, \ldots, S-1)$ denote

Can truly replicate and better build on each others' work if we emulate the genomics community and require algorithm and code deposition

Graduate student training in our core science areas clearly requires explicit (rather than implicit) training in computational best practices



Take Homes

- Probability underutilized
- Crowds good for more than clouds
- Time series rock
- 900 m^2 often too big
- Pixels have neighbors
- Algorithms and computation essential to our science
- (and, way, don't mess with success)